DOCUMENT RESUME

ED 319 604 SE 051 414

AUTHOR McKnight, Curtis C.; And Others
TITLE Task Analyses of Critical Evaluations of Quantitative Arguments: First Steps in Critical Interpretation of Graphically Presented Data.
PUB DATE 90
NOTE 23p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Case Studies; *Cognitive Processes; Cognitive Psychology; *College Mathematics; *Data Interpretation; Data Processing; *Graphs; Higher Education; *Information Processing; Interviews; *Mathematics Skills; Protocol Analysis

ABSTRACT

The ability to think critically in the presence of arguments with essential quantitative elements, most often graphical elements, will become an essential skill for educated citizens in the future. This paper takes one specific graphical display, a narrow series of observational and interpretational tasks related to a graph, and using a small set of case studies, takes some initial steps in characterizing the organization and nature of cognitive information processing in relation to those tasks. Seventeen college freshmen were interviewed using a graph and three questions on the graph; data from 10 subjects was analyzed. A model of data point reading; the organization of cognitive processing; and a catalog of processing errors and off-task processing were discussed. Some distinctions in the processing are outlined for further exploration. (YP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED319604

Task Analyses of Critical Evaluations of Quantitative Arguments: First Steps in Critical Interpretation of Graphically Presented Data

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Curtis C. McKnight

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Curtis C. McKnight
Corinne Kallman
Mark Fisher

University of Oklahoma

Paper presented at the annual meeting of the American Educational Research Association, Boston, April 1990

SE051414

BEST COPY AVAILABLE

The entrance of our society into the information age, with its ubiquity of computer-enhanced publishing and presentation graphics, has led to a virtual bombardment of both citizen and student with numerical data and, moreover, with numerical data embedded not in the context of separate treatises or tests on mathematics, but in the context of informative articles which often contain either central or peripheral arguments that have essential quantitative elements. In particular, there has been a marked increase in presentation of data in graphical form, whether used to further the discussion to which it is attached or used, more peripherally, for cosmetic purposes in the presentation of textual material. A variety of uses have been suggested (e.g., Lefferts, 1981) for such graphics: clarifying, simplifying, emphasizing, summarizing, reinforcing, arousing interest, providing impact, increasing credibility, and enhancing coherence. To this catalog of legitimate uses of graphics and numerical data must, unfortunately, be added their perversion in arousing interest without providing appropriate substance, serving the cause of credibility when such credibility is not merited, producing impact sufficient to overcome innate critical responses, and so on.

The ability to think critically in the presence of arguments with essential quantitative elements, most often graphical elements, will become (if it has not already become) an essential skill for educated citizens in our society. Instruction related to the skills necessary for such quantitatively oriented critical thinking will certainly enter the curricula of school mathematics if the new National Council for Teachers of Mathematics (NCTM, 1989) standards for mathematics as reasoning and communication have their desired impact.

In the face of the current impact of critical evaluation of quantitative and graphical data in arguments, and the likely growth of this impact in the next few years, surprisingly little direct research has been done on the cognitive processing that is involved in successful information processing for these sorts of tasks. Obviously, there are extensive research literatures on problem solving generally and in mathematics more specifically. Further, there is also an extensive literature on reasoning in general (e.g., Glasman, Koff & Spiers, 1984; Siegler, 1978), even to the point of generating undergraduate texts for critical thinking (e.g., Halpern, 1989).

Direct research into information processing and critical reasoning about graphically presented data is considerably more sparse. Early work was done on interpreting graphically presented information by Washburne (1927) and Vernon (1946, 1950). Several studies also investigated graphical information as a part of investigations related to audio-visual communications (Peterson and Schramm, 1954; Culbertson and Powers, 1959; and MacDonald - Ross, 1977). Two more recent works on graphical presentation of information contained some research results (Tufte, 1983; Cleveland, 1985). Finally, there have recently been a small number of research studies investigating graphical comprehension and reasoning in relation to graphing displays (Eggen, Kauchuck and Kirk, 1978; McKnight, in press) or comprehension of relationships in mathematical graphs (Curcio, 1987). In view of the importance of the topic, it seems clear that additional research is needed. Indeed, other sessions at the present A.E.R.A. meeting suggests that these issues are again attracting attention, at least at the level of graphical display of information in statistics and in comprehending information presented on computers.

The present paper reports on on-going work, a part of a more general research agenda discussed elsewhere (McKnight, in press). The present paper takes one specific graphical display, a narrow series of observational and interpretational task related to that graph, and, through a look at a small set of case studies, tries to take some initial steps in characterizing the organization and nature of cognitive information processing in relation to those tasks.

The Experimental Situation

Learning to read graphs, especially something as simple as graphs with several points plotted but unconnected by lines, is a set of simple skills learned early in schooling and certainly before entering college. While the amount of time devoted to remedial mathematics in college suggests that the above statement is an oversimplification, essentially it is correct and represents a set of skills assumed – or, at most, reviewed – in college mathematics courses.

Transforming that same set of skills into a graphical situation in which data have substantive meaning – the typical situation of presentation graphics used as a simple form of descriptive statistics – would seem to be at best a small extension of the application of that same set of graph reading skills. Such is not the case. The introduction of substantive meanings to the elements of simple graphs – and supplementing skills at reading graphs with observing and interpreting graphical patterns in terms of those substantive meanings – turn out to involve a range of new, differing and complex skills beyond those of simple graph reading. The new, more complex array of interpretational and observational skills introduce new demands for organizing cognitive processing tasks effectively. Further, these additional complexities introduce whole new classes of confusions and error patterns when student work is examined.

Modeling the cognitive processes of graph comprehension and interpretation, the organization of such processing and the errors endemic to these processing tasks will require extended and detailed investigation. What is attempted here are a few initial steps to provide a "rough map of the terrain" to guide further work. The experimental task that was used will first be described, followed by sketches of some early results from this investigation. Further analyses of the data and of a greater number of cases are on-going.

The Instrument

An early item on a research agenda devoted to understanding cognitive processing of quantitative arguments must be the task of understanding more clearly the skills involved in reading graphs and, in particular, in reading and interpreting presentation graphics – graphs presenting substantive data about the real world for which axes, units, amounts, etc., have meaningful interpretations. This paper focuses on one single graph, taken from an article in *Scientific American* (Cohen, 1987), and on three of a longer series of questions asked about this graph. Figure 1 reproduces a simplified adaptation of the graph and Table 1 the three task-setting questions used for this study.

Insert Figure 1 and Table 1 here

The Subjects and Procedure

Seventeen college freshmen were interviewed during the first week of a seminar for freshmen on "The Rhetoric of Persuasion in Statistics". Since the seminar met weekly, this meant that each student was interviewed after the first, organizational meeting but before the second meeting in which substantive issues of the critical evaluation of graphical data were addressed.

Each student answered the three questions seen in Table 1, among other questions dealing with this graph and with three other graphs, talking aloud about what they were doing and what they were thinking about as they did so. These sessions were audiotaped. The initial question-answering was followed by a "retrospective" interview in which the researcher sought to clarify points of uncertainty in the original answers by asking students to expand on their answers and thinking processes. Ericsson and Simon (1984) point out the differences between such direct talk-aloud data and retrospective talk aloud data. Both types of data were gathered in this investigation and the distinctions between them were considered carefully in data interpretation.

Transcriptions were made of these interviews and they are being analyzed and coded by various schemes. These data will eventually be used to form information-processing models of the skills students use to answer such questions, models of the organization of those procedural skills and detailed descriptions of the failures in those processes that generate various error types. Some preliminary steps in this direction are reported below. For the present paper, transcripts for ten subjects were selected at random for more detailed analysis.

Towards a Task Analysis of Data Point Reading

Of the many tasks of interpreting graphical displays of data, one of the simplest is the location of a specified point on the graph, estimation of its value according to a specified scale of the graph, and interpretation of that value in terms of the units of that scale. To set such a task requires a graph in which axes are labeled with numerical scales and in which the substantive variables involved and their units are labeled. Further, it requires some form of point labeling, at least for the point in question, so that the point can be referred to in task instructions without stating its data value or values. In addition to this sort of graphical stimulus, a set of task instructions must be provided which identify the point to be located, identify the axis the scale of which is to be used in estimating the value of the variable for the given point, and make clear whether or not the value of the specified point for the specified variable is to be stated in terms of the units of the scale or as a unitless numerical value (the unit being typically included in the question statement).

The graph in Figure 1 certainly fits the criteria for the type of graphical stimulus needed. In the actual graph in Cohen (1987), 39 points were presented in the graph with a country label for each point. The simplified adaptation in Figure 1 uses a smaller number of points

and a few country labels to identify the more complex stimulus shown to subjects. A heading was provided for the graph, "Diet and Cancer," and each axis was labeled with a variable name and unit — "Age-adjusted deaths per 100,000 population" for the vertical axis and "Estimated fat intake (Grams per day)" for the horizontal axis. In addition, the following caption was presented as part of the stimulus: "STRONG LINEAR CORRELATION is seen when data for dietary fat and deaths from breast cancer are plotted. (Fat intake was derived by dividing total fat consumed in a country by population; waste and consumption by animals were not taken into account.)."

The task instructions and the set of choices for the multiple-choice task (Question 1) fit the criteria for setting the desired type of task. A specific point was identified by using its country label, "Hong Kong." A variable and hence an axis were identified ("Approximately how many (age-adjusted) deaths ...") for which an approximate value of the given data point was to be determined. Since the question was phrased in terms of the unit (a rate — "deaths ... per 100,000"), it was implied that the answer should be stated as a value from the scale of the graph along with an appropriate unit . The wording of the choices reinforced this. The question was, however, complicated by the fact that the rate unit involved a scaling factor ("per 100,000") which was not to be included in the answer statement since it was excluded by the question and not stated as part of the choices for answers. Note also that only in the title for the graph was mention made of "cancer" and only in the caption was the data linked to "breast cancer" in particular.

A Model of Data Point Reading

Without attempting to further analyze cognitive processes into smaller units, the following procedures (or, if preferred, linked chunks of cognitive information processes) certainly seem sufficient for successfully completing the processing task needed to answer Question 1 (other than by guessing):

1. **Identification of point specified** (Possible sub-processes: identify task as point value estimation task, identify sub-task of point identification, understand country name as point label [correspondence of points and labels], locate country name in graph field, parse task instructions for country label, identify point corresponding to country name located in graph field, etc.)
2. **Identification of axis from which value is to be read** (Possible sub-processes: identify sub-task of axis identification, understand axis-legend connection, parse legend for possible axis identifying words and phrases, parse task instructions for axis-identifying word or phrase, match corresponding identifiers to locate specified axis, etc.)
3. **Estimate numerical value corresponding to point and axis** (Possible sub-processes: visualize line from point orthogonal to identified axis, identify approximate interception of visualized line and identified axis, find number corresponding to approximate interception treating axis as number line [scale indicated by cross-hatches and indicated numerical values], etc.)
4. **Identify units of desired answer** (Possible sub-processes: parse question statement for unit identifiers, parse answer choices for unit identifiers, match question and answer-choice unit identifiers for congruence, determine desired unit by selection of appropriate

unit label, etc.)

5. **Convert number in estimated value units to number in desired units if differing** (Possible sub-processes: parse axis legend for unit identifiers; identify, from axis unit identifiers, unit in which estimated numerical value is expressed; decide if further processing is needed to match units of numerical value as estimated to desired unit; initiate further computational sub-procedures for unit conversion as needed; etc.)
6. **State answer choice** (Possible sub-procedures, multiple-choice context: compare estimated or computed numerical value with numerical value in each answer choice; if match, indicate letter of choice identifying match; if no match, indicate letter of choice indicating unable to answer from given data [in this case, graph], etc.)

Clearly, this set of processes and this ordering of processes is sufficient to produce a correct answer in this case and forms one model of how this information-processing task might be accomplished to obtain a correct result. A more complete model obviously would include a default to successful guessing at any point in the processing performance. Further, there is nothing canonical about the order of these suggested procedures. Clearly the first and second procedures suggested might be permuted (as could be many of the possible sub-procedures indicated). Also, the procedures, and certainly the sub-procedures, listed are sufficient, not necessary. Although no evidence to the contrary was found in these case studies, it is possible that a more detailed study or a greater number of cases might reveal different procedures than those set out above. Little explicit evidence can be adduced from the case studies for the possible sub-procedures and additional detailed investigations might easily suggest differences here.

Finally, in the organizing of individual processing performances, there is no guarantee (but no explicit evidence to the contrary here) that sub-procedures from the same procedures must all be accomplished before passing to the next procedure. That is, the organization of sub-procedures into larger procedural chunks is idealized. Actual performances, especially non-expert subjects may often consist of mixed sequences of sub-procedures from different procedures, of repetitions, of checking by re-processing, of omissions and procedures for identifying omissions, etc. The "search path" of procedures from beginning state in perceiving the stimulus to end state in indicating an answer choice is likely to be quite complex for many individuals and may often include false starts, dead ends, loops, etc.

Some Cases of Actual Data Point Reading

The answer choices set out in Question 1 were designed to provide some indication of the procedures used by subjects and, in particular, which procedures fail. [No correction, of course, was possible in this investigation for the default to guessing, either correctly or incorrectly, other than explicit statements in verbal reports.] The correct choice, "a", is at least a rough indication that all procedures likely functioned correctly. Incorrect choice "b" is most easily explained by the failure of either Procedure 4 or 5, resulting in an inappropriate or inaccurate scaling of a read value to an answer value in differing units. Incorrect choice "c" seems to indicate a failure in Procedure 2 resulting in further processing using the inappropriate axis. Incorrect choice "d" would seem to indicate some combination of a

failure of Procedures 2, 4 and 5, or guessing. The "non-answer" choice, "e", provides no information by its selection other than a failure to successfully complete the task. Protocol analysis is needed to identify the specific failure that led to selection of this choice.

Of the ten subjects whose transcripts were used for the present study, six made the correct choice ("a"), three selected the incorrect choice "b", and one selected the incorrect choice "e". In the case of the single subject who chose "e", it was clear from his transcript that the failure was due to not locating the point corresponding to Hong Kong. In the initial talk-aloud portion he stated, "... Can't tell from the graph, 'cause Hong Kong is a city and these are all countries." When questioned further as indicated in the retrospect section of the transcript he stated, "Well, I couldn't find Hong Kong on there, so I figured that you couldn't tell from it." On further prompting that Hong Kong was indeed on the graph, he located it and arrived at a correct answer. Clearly the initial failure was of Procedure 1, locating the specified point. What is not clear is whether he searched and could not find the point initially and used his city/country distinction as a justification for that failure or, if, after minimal (if any) searching, he arrived at the city/country distinction and from it concluded that this was the reason for his failure and that there was no need to search further.

No subjects chose either "c" or "d" indicating that it was likely that there were no difficulties with Procedure 2, the identification of the desired axis and, indeed, the ten transcripts showed no evidence of that sort of difficulty. Three subjects chose "b" and, in each case, the confusion was on interpreting the rate label ("deaths per 100,000") as a scaling factor (i.e., the desired units were to be scaled as the "actual" number of deaths, not the number of "100,000's"), hence calling for a multiplication to re-scale to the desired units. The answer choices "a" and "b" made the answer possibilities neutral with respect to this decision. The question asks for "deaths ... *per 100,000*" [emphasis added] and thus called for no re-scaling. The failure thus seems to be in Procedure 4, identifying the desired units, which resulted in an inappropriate use of Procedure 5. A transcript section from a subject choosing "b", typical of the other two who made this mistake but somewhat more explicit about the processing, sketches the subjects processing as follows:

Student: "... Working on [Question 1]. I'm trying to find Hong Kong on the graph. ... I found it. ... at ... one million deaths. [Choice]B, because I followed the line over to age-adjusted deaths per 100,000 and found it to be near 10, 10 times 100,000 makes a million deaths is the way I came up with the answer."

Within the six transcripts of students who successfully completed the task, explicit evidence was found for all of the procedures outlined earlier (but not for all of the sub-procedures suggested as possible). Even the transcript section quoted above for the student who was unsuccessful (chose "e") suggests the operation of most of the procedures postulated ("trying to find Hong Kong on the graph", "I followed the line over to age-adjusted deaths", "near 10", etc.). The empirical evidence from the ten transcripts do nothing to contradict at least the broad outlines of the model for data point reading postulated above. It should be recalled, however, that it is likely that the actual flow of sub-procedures is more complex

than suggested by the initial model although that cannot easily be verified by the use of protocol data and certainly not from a sample this small.

The Organization of Processing

Certainly the detailed modeling of cognitive process sequences in completing information-processing tasks relative to "real world" data presented in graphical form is both useful and important. The example above suggests a start for one such task and some of the methodological and empirical questions raised. Of perhaps even more interest than the modeling of many processing procedures or skills oriented to specific tasks is the related question of how information processing is organized. Suggestions have already been made of the likely complexity of processing at the flow of sub-procedure level. Still other phases of processing have yet to be discussed.

When a situation is created in which several information processing tasks are set for a subject relative to the same single graphical stimulus, the possibility is created for an even more complex flow of cognitive processing. When tasks move from "simple" observations of isolated graph elements (e.g., Question 1 in Table 1) to observation of patterns (Question 2) or the interpretation of patterns in substantive terms of the "real world" data presented (Question 3), considerably more potential for complexity in the flow of processing is created. This is accompanied also by the greater complexity of the move from selecting responses (Question 1; Question 2, choices a–c) to generating responses on one's own (Question 2, choice d; Question 3).

In this context, complex episodes of cognitive processing flow from the single stimulus and the sequence of task. Additional complexity is introduced by retrospective questioning which encourages re-consideration of answers to earlier questions after all questions have been answered. These complex episodes make possible a variety of new questions relative to the more global features (as opposed to the more local, sub-procedural features) of processing organization. Some questions of this sort include:

1. Which features of the stimulus are heeded?
2. What drives the processing within an episode? Is it driven by the attempt to understand the graphical stimulus which would allow completion of multiple tasks or is it driven by the specific questions setting the information-processing tasks?
3. Is the initial consideration of the graph, whether question-driven or not, focused narrowly on specific points and details of the graphical stimulus or is it more widely focused on the graph as a whole or, at least, on patterns and subsets of points?
4. At what point does the effort at interpretation — at the translation of statements about mathematical facts, relationships and patterns into statements about the substantive, "real world" meaning behind those things — begin? Does it occur spontaneously and early in a processing episode as an example of an attempt as sense-making, or does it occur relatively late in the processing episode, most often stimulated by questions which set tasks which can be completed only by interpretation?
5. Given a graphical stimulus which provides a limited set of "real world" information —

certain facts and relationships, assuming the truth of the data graphically presented — does the processing that results restrict itself to the information given or does it seek to go beyond that given (which was adequate for the task) to include cognitively-linked additional information (or, perhaps, misinformation) about the "real world"?

Certainly data from ten subjects cannot resolve these questions in any definitive way. However, they can serve to show that they are indeed real questions and point to facets of the organization of cognitive processing that are empirically problematic in understanding the handling of graphically presented data. In short, the limited data available here are sufficient to point to some real distinctions in need of further exploration. Each of these will be briefly considered.

Heeded versus Unheeded Stimulus Features in Processing

Which features of the stimulus are heeded? The graphical stimulus presented in the present study was complex and offered a number of features that could be attended to or ignored by subjects. While all elements of the stimulus graph were likely in the subjects' perceptual fields at some point, it is not reasonable to assume *a priori* that they attended to all features, let alone that they divided attention among all features evenly. It is possible that a small set of features dominated in all cases, or that a small set of features dominated in individual cases but differed from case to case. The salience of kinds of features for cognitive processing of graph tasks is obviously an essential target in further investigations but is difficult to investigate by talk-aloud methods with a single stimuli. Some direct verbal report information can come from transcripts but that provides no information on dominance of a feature, nor does lack of explicit verbal report disprove heeding a feature.

In the case of the present study, the pool of data from the ten subjects indicated that virtually all aspects of the graphical stimulus were heeded by one or more subjects: the stimulus title, the figure caption, both scale legends, and point labels. In addition, there was direct evidence that one or more subjects considered the scale values of the selected point (and in some cases other points) according to one or both axes, the location of points broadly (upper left, etc.), the number of points, specific subsets of points (e.g., lower right points identified with developing countries), and the pattern of points ("The points seemed to fall mostly in a line ..."). While this is a fairly complete list of possible features, considering only direct verbal report as evidence, no subject appeared to attend to all features (although this may well be an artifact of the limitations of verbal report data). Each spoke about a relatively small number of features and the features mentioned differed from subject to subject. Further, more fine-grained investigation is in order.

Graph-driven versus Question-driven Processing

What drives the processing within an episode? Processing might be driven only by the specific questions that set information processing tasks. That is, a subject might begin explicit (in terms of the verbal report data) consideration of the graph stimulus in detail only after processing at least one task-setting question, Question 1 here. By contrast, processing might be driven by the knowledge of the general task environment — several processing tasks using the same graphical stimulus — and be focused initially on the features of the

stimulus, on understanding and relating them, before direct attention was turned to any task-setting question. This distinction was captured by deriving operational definitions from the protocol analyses that allowed most subjects to be classified as either "question-driven" or "graph-driven".

Only eight of the ten subjects could be unambiguously classified into these two categories. Two were so terse in the initial phase of their talk-aloud performance that they could not be classified. Of the eight who could be unambiguously classified, four fell into each category. This suggests that the distinction is a useful one and that there is an empirical difference in processing organization tapped by this distinction. Neither category was systematically related to quality of processing performance, some more and some less successful performances falling into each category. There was also no systematic relation between these categories and the small amount of background information available (gender, age, high school GPA, and Math ACT score). Thus, differences here seemed not to be a function of academic ability or past achievement, nor even differentially effective in processing success. Whether they constituted personal processing-organization styles is an open question although there is some evidence that this may be so. All subjects interviewed were actually exposed to tasks related to four differing graphical stimuli and, while all of that data has not yet been analyzed, there appears to be some consistency across processing tasks. Certainly exploration of this distinction in organization, its consistency, and its basis merits further investigation.

Wide-focused versus Narrow-focused Processing

The initial consideration of the graph, whether question-driven or not, could focus narrowly on specific points and details of the graphical stimulus or could be more widely focused on the graph as a whole or, at least, on patterns and subsets of points. Again, operational definitions were developed to classify subjects into "wide-focused" versus "narrow-focused" processors on the basis of verbal report data. Again two subjects could not be unambiguously classified (one of whom was previously unclassifiable as question- vs. graph-driven). Of those classifiable, five were wide-focused and three were narrow-focused. Again there was no systematic basis for these differences found in background data, nor were they systematically related to effectiveness. Again there seemed, preliminarily, to be some consistency across processing tasks. Comparing this distinction to that of graph- versus question-driven processing, there was systematic relationship, at least one subject falling into each of the four combinations possible for these two distinctions. Further investigation again seems in order.

Interpretation-oriented versus Task-oriented Processing

At what point does the effort at interpretation — at the translation of statements about mathematical facts, relationships and patterns into statements about the substantive, "real world" meaning behind those things — begin? Does it occur spontaneously and early in a processing episode as an example of an attempt as sense-making, or does it occur relatively late in the processing episode, most often stimulated by questions which set tasks which can be completed only by interpretation? Certainly these questions are related to a difference of potential salience, whether it is related to differential effectiveness or merely indicates

differing processing styles (and is related, perhaps, to personality or other cognitive style features).

This distinction was again captured by a set of operational definitions to use verbal report data to classify subjects as "interpretation-oriented" or "task-oriented". The former began interpretive, translational moves in a graph "preprocessing" (prior to engaging specific tasks) phase or in relation to Question 1 which called only for observation rather than interpretation. The latter began interpretive moves only on engaging Questions 2 or 3 which had more explicit demands for interpretation (indeed, Question 3 is a purely interpretive task). Again two subjects (not the same as before) were not unambiguously classifiable, since some remarks in preprocessing or in answer Question 1 might be construed as interpretive, sense-making efforts but not unequivocably so. Of those who were classified, four fell into each of the two categories. Again there was no systematic relationship with background features or with processing success.

Unlike previous cases, however, there was at least one relationship with the processing-organization distinction previously discussed that seemed to merit further investigation. Every interpretation-oriented subject was also graph-driven (four subjects) and every task-oriented subject was also question-driven (three subjects). While the numbers are insufficient to guarantee that this result is not an artifact of the experimental situation, it seems unlikely in that for some of the interpretation-oriented subjects, even though there was a graph-driven "preprocessing" phase, the interpretive moves did not always occur in that phase, but for at least two subjects occured after Question 1 was engaged. Certainly those who were graph-driven would seem, *a priori* to be more oriented to sense-making and more likely to be interpretation-oriented, but this relationship was not necessary as an artifact of how the distinctions were defined. Clearly this relationship is interesting and worth further investigation.

Externally-connected versus Internally-connected Processing

Given a graphical stimulus which provides a limited set of "real world" information — certain facts and relationships, assuming the truth of the data graphically presented — processing could be restricted to the substantive information explicitly given or it could seek to go beyond that given (and adequately) for the task to include cognitively-linked additional information (or, perhaps, misinformation) about the "real world". Again, operational definitions were derived to classify subjects into "internally connected" processors (those who worked only with the information given in accomplishing the information processing tasks set) and "externally connected" processors (those who went beyond the information given [e.g., to bring in information known or assumed about specific countries, their diet, etc.]). No use was made in defining the categories of whether or not the assumed "information" brought into the task was correct, nor of whether or not the linked to "external" information was more cognitively oriented ("Thailand is a less developed country.") or affectively-oriented ("I saw South Africa was way over there [upper left] and it kind of irked me 'cause I know a lot of black people are getting killed in South Africa and that has nothing to do with how much fat they eat.")

One subject could not be unambiguously classified, three were internally-connected pro-

cessors (restricting themselves to the information given) and five were externally-connected processors (going beyond the information given). There was a slight linkage between internally-connected processors and success in processing because, undistracted by other presumed relevant information about countries involved, they were more likely to focus on the relationship between the two factors in the stimulus (mortality rate from breast cancer and dietary fat) and to look at the pattern of dots in the scatterplot rather than imposing a pattern based on other distinctions (more versus less developed countries). Otherwise, there was no systematic relationship between categorization by this distinction and by any of the other distinctions or background factors. Certainly, however, going beyond the information given did make subjects more prone to certain processing "errors" (or, at least, off-task processing), even when they were eventually successful in the processing tasks. Some of this can be seen below in the next section.

A Catalog of Processing Errors and Off-task Processing

Certain types of graphical task processing leads to erroneous answers and such processing can fairly be labeled erroneous. Other talk-aloud protocol transcripts reveal something that might better be considered "off-task" processing, since it leads to false starts and dead ends in terms of the goals of the processing tasks, but may not prevent the processor from eventually arriving at those goals and answering questions correctly. Such off-task processing can be considered an "error" only in the sense of not being a part of the most direct route to the processing goal. Nonetheless, for simplicity, both categories of processing will be considered here as "processing errors" broadly construed.

A part of any future research agenda for investigation in the broad area of the present study should include extending the catalog of processing errors and explicating the genesis and nature of those errors, since that would seem a necessary key to developing effective interventions to improve critical thinking in relation to graphically presented data. Since work is at a fairly early stage in modeling the cognitive processing of graphical information processing tasks, it would be presumptious to make anything other than general remarks about the genesis and nature of processing errors. However, at least a preliminary catalog of such errors can be begun and that, on the basis of the transcript data for the ten subjects of this study, is done here.

General Confusion

One of the first reactions — and a reaction that, until overcome, prohibits useful processing of the information presented in a graph — is that of overwhelming confusion by the details of that graph. One student described his reactions, upon first examining the graph in Figure 1, as follows:

> "The graph was kind of confusing at first ... There was a lot of little dots everywhere and then once I started reading what was near the dots, and then reading under the graph, I understood what they were trying to ask for ... That's what my first impression of the graph was."

This student spent considerable time before answer the first question and, at that, missed even that first question. By comparison with presentation graphics (with substantive variables for each axis, often with complex units for those variables [as here a rate per 100,000], legends or data point identifiers, etc.), abstract, mathematical graphs are much cleaner and more straightforward. In mathematical graphs, general confusion about the complexity of the information arrayed rarely arise. This suggests that direct research on these "messier" graphs of empirical data are needed, along with research on interpretive information tasks made possible by this sort of graph.

Specific Misreadings

In addition to general confusion, the complexities of presentation graphics often lead to specific misreadings. For example, with the graph in Figure 1, the variable for the ordinate was "age-adjusted deaths per 100,000". That variable thus had both a substantive meaning (a mortality rate, with the further complexity of being age-adjusted – a fact unexplained in the graph as presented) and also a scale factor for the variable, in this case a *rate* as indicated by "per 100,000". Thus, even if the correct value, 10, was read from the graph for Hong Kong in answering Question 1, it was still possible to misinterpret the unit and arrive at an incorrect choice. For instance, several students made the error detailed in the following fragment of verbal data:

> "Working on [Question] 1 ... I m trying to find Hong Kong on the graph ... I found it.... At one million deaths the answer is B, because I followed the line over to the age-adjusted deaths per one hundred thousand and found it to be near ten. Ten times one hundred thousand makes a million deaths is the way I came up with the answer."

Thus, even when this student overcame his initial confusion and tried to answer a straightforward graph reading question, the additional complexity of the substantive situation — in particular, misinterpreting the desired unit, which initiated an inappropriate scale correction — led him to an incorrect answer. This kind of specific error can actually be characterized at a reasonable first level of detail by the model presented in the first section of this paper.

Prototypical Graphs – The Way It's S'pozed to Be

Students are often handicapped in these tasks by the very skills and expectations they have inherited from previous experience with graphs, especially (abstract) mathematical graphs. A typical example is provided by a student who had clearly in mind what this sort of graph should be and felt unable to provide adequate answers to the questions because this graph didn't fit that prototype:

Interviewer: [after initial answering period] "Did you have any trouble with this graph?"

Student: "Just in the beginning.... I think that it needs bars going all the way up, and bars going all the way across.... It's just a bunch of dots. Then you have to search and try

to find the right number."

One area for further research should be the exploration of mental models of prototypical graphs. This would both provide a basis for better characterization of cognitive processing in relation to graphs (processing defaults, etc.) and also make possible a clearer understanding of certain types of errors.

Distraction by Extraneous "Real World" Knowledge

Translation from the "clean" world of abstract mathematics to the "messy" world of everyday reality — in which all of our knowledge has links to other knowledge as well as links to personal beliefs and emotional reactions — introduces yet another complexity. Sometimes that other knowledge — or what one thinks is other relevant, linked knowledge — or those beliefs and affective reactions interrupt the more cognitive, information-processing tasks of interpreting the graph. Good examples are provided in the case of some students' responses to the labels for various countries in the graph of Figure 1, responses that integrated their beliefs about those countries into the task of answering questions about the patterns that are there. Two examples are:

First example:

Student A: "Well... Thailand, Philippines, Japan, Mexico, Columbo *[sic]*, and El Salvador are among the non-prosperous nations of the world.... New Zealand, Switzerland, USA, Germany...." as writing, for [Question] 3, the statement "The underprivileged nations of the world (listed in the bottom corner) have fewer fat people and age related deaths then the more urban and economical prospores nation of the world." *[sic]*

Second example:

Interviewer: "How did you feel when you looked at that graph [that for Figure 1]?"

Student B: "I don't know. I saw South Africa was way over there and it kind of irked me, 'cause I know a lot of black people are getting killed in South Africa, and that has nothing to do with how much fat they eat.... That's one of the things I saw first."

Student A's inclusion of Japan among the "non-prosperous" nations allowed him to go on with his over-generalization, rather than focusing on dietary differences that had little to do with prosperity and more with cultural patterns. Student B's reaction to South Africa — which triggered strong affective reactions for her — kept her at first from even focusing on the task at hand and beginning to examine the other aspects of the graph, even though she admitted that the deaths she was thinking of had nothing to do with dietary fat and were thus not relevant information for interpreting the graph.

Multiple Focus in Interpreting Graphical Patterns

Not only does the inclusion of real-world information sometimes lead to non-relevant belief linkages or affective reactions, it also makes possible more than one focus in interpreting the patterns that may or may not emerge from the data presented in the graph. This is part of what makes salient the distinction between internally- and externally-connected processors. For the graph in Figure 1, the inclusion of country names made it possible to focus on differences among countries (e.g., relative wealth, third-world status, etc.) as well as on the differences specifically and directly portrayed in the graph — dietary fat intake and breast cancer death rates. This possibility of different foci led to differing types of generalizations as may be seen in Table 2, which presents the answers given by several students to Question 3. The transcripts illuminate some of these differences further:

First example:

Interviewer: "What did you think when you saw this graph?"

Student A: "After I read through it, I thought it was a pretty simple graph.... They have quite a bit of information on it with all the different countries, but it really does show it well.... You can see that the dots pretty well go in a line." Produced for Question 3: "As fat intake increases, so do the number of deaths due to cancer." [a focus on the information presented directly in the graph]

Second example:

Student B: "The dots on the bottom of the graph are third world countries, and the higher dots are mostly European countries. And the European countries, they take in more fats than the third world countries.... There really is no pattern, except that the lower countries with the lower death rate are really like third world countries. Not as developed as UK, Switzerland. And the ones on top, they're mostly European countries.... The ones on bottom, ... there's some from Central America. And Philippines. Japan, that's in Asia. So, like the lower countries are third world countries and the higher ones are more advanced." [an attempt to use country labels to link to other information and, using an imposed categorization, derive another pattern less directly linked to the information presented in the graph]

Third example:

Interviewer: "Why did you choose letter C for your answer [to Question 2]?"

Student C: "Well, I was thinking of it in a country sort of way, that the countries don't make it [Query: What is 'it'? The pattern of the dots?] relevant information....

[Continuing in the retrospect, after the initial answering phase:]

Interviewer: "Would you still give the same answer?"

Student C: "... I would give a different answer and say that more fat intake would be the higher population, the death would be higher." [after a clarification by the interviewer, the student indicates this as answer A to Question 2; this is an example of a student who first focused on the country labels and saw no pattern then, on prompting, focused on the presented data and saw a pattern that led to the correct choice]

Conceptual Errors in Interpreting Scatterplots

A final category of error is that of conceptual errors in interpreting scatterplots, e.g., inferring cause and effect from correlation. One error of this kind is perhaps subtle, but virtually universal with freshmen students. The dots of the sort of scatterplot seen in Figure 1, even if they suggest a positive linear correlation, not only do not imply cause and effect (a mistake largely not made even by these freshmen), but also do not refer to the same subject at different moments in time. It is cross-sectional rather than longitudinal. Thus the appropriate language for interpreting Figure 1 is associative, as in choice A for Question 2, rather than dynamic as in "As more fat is taken in, the rate of breast cancer increases." The latter sort of answer embeds the dynamics of moving across the graph into the dynamics involved in the answer, as if the same object was being tracked over time and through various changes.

Insert Table 2 here

Table 2 presents the written responses for the ten subjects in answer to Question 3. Some of the responses indicate the previous error of bringing in non-salient real-world information — see Responses 7 and 8. Response 4, when examined in the context of the fuller context reveals an initial failure because an attempt was made to use such external information about countries but without sufficient knowledge to find the pattern. Response 3 is more a statement of the fact that there is a relationship rather than a statement of the relationship; Response 6 also seems to fall into this category. Cause and effect language is seen in Responses 2 and 10, although this confusion seemed relatively rare even with freshmen students (one wonders, of course, given Response 2, whether it is the fat that causes death or the estimation process — perhaps this is a phenomenon which should be investigated further by mathematics education researchers).

As can be seen from Responses 1, 5, 9 and 10 in Table 2, dynamic language was often generated by students in interpreting the graph's pattern as an answer to Question 3, when other sorts of errors were not made. Associative language was used only when already given in the choices of a multiple choice question [as Question 2]. The students' noticed no difference and felt no conflict. Further, there seemed also to be a willingness to move from a graph which presented information at the country level to conclusions stated at the

individual level. This can be seen in the "your" of Response 5. It can also be seen by a more detailed look at some transcripts, e.g., that of the student who generated Response 9 in Table 2:

Student: [initial response to Question 1] "OK. It looks like I'm going to start [the graph in Figure 1]. Deaths from breast cancer plotted [against?] age-adjusted deaths per hundred thousand. What this is trying to tell me is that in Thailand they had approximately .7 age-adjusted deaths per hundred thousand population and they ate 28 grams of fat per day. It's kind of giving me two different things. Whether they're related or not I don't know. That makes sense." [Answer to Question 3] " The graph indicates that as fat intake per capita increased, the mortality rate in the individual populations from breast cancer increased as well."

Segment of retrospect for same subject:

Interviewer: "Did you have any problems with [Question 1]?"

Student: "... I thought I knew pretty much what the graph was trying to tell me...." [After some further probing:] "The people that died were probably the ones that ate a lot of fat. You know. Maybe the people that got breast cancer were the ones that ate a whole lot of fat...."

Interestingly, both this use of associative language and the move from the global to the personal was often true for other, more expert [advanced graduate students, faculty] subjects who took part in a related experiment. While the latter was suggested by the very nature of the graphical stimulus, it certainly was not supported without further evidence and the data of Figure 1 could at best be considered suggestive of the truth of the relationship at the individual level. This points to a fundamental principle that needs to be brought ought in education about descriptive statistics and statistical concepts, rather than a simple error pattern in graph interpretation skills.

Concluding Remarks

The investigations presented above were clearly preliminary in nature — only one stimulus (and related tasks) was used, data from only ten subjects was analyzed, and only some of that analysis was pursued in detail. However, even this broad sketch of preliminary results seems enough to suggest the value of further, detailed investigation of comprehension of, interpretation of, and critical thinking about data presented in graphical form. Some specific suggestions for further investigation were presented throughout the paper but the general agenda surely must include characterizing specific processing procedures, investigating the nature and basis for processing organization, and cataloging processing errors and exploring their genesis and nature. As the look at the literature suggests, this research domain has large gaps. As a look at the front page of your local paper or an issue of *Time* magazine suggests, there is a marked need to fill those gaps and quickly.

References

Cleveland, W. S. (1985). *The elements of graphing data.* Monterey, CA: Wadsworth Advanced Books.

Cohen, L. A. (1987). Diet and cancer. *Scientific American, 257(5)*, 42-48.

Culbertson, H. M., & Powers, R. D. (1959). A study of graph comprehension difficulties. *Audio-Visual Communication Review 7(1)*, 97-110.

Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education 18(5)*, 382-393.

Eggen, P., Kauchuch, D. & Kirk, S. (1978). The effects of generalizations as cues on the learning of informations from graphs. *Journal of Educational Research 71(4)*, 211-213.

Ericcson, K. A., & Simon, H. A. (1984). *Proctocol analysis: Verbal reports as data.* Cambridge, MA: The MIT Press.

Glasman, N. S., Koff, R. H., & Spiers, H. (1984). Preface [to special issue on teaching and learning critical thinking]. *Review of Educational Research, 54(4)*, 461-471.

Halpern, D. F. (1989). *Thought and knowledge: An introduction to critical thinking.* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lefferts, R. (1981). *Elements of graphics: How to prepare charts and graphs for effective reports.* New York: Harper and Row.

MacDonald-Ross, M. (1977). How numbers are shown: A review of research in the presentation of quantitative data in texts. *Audio-Visual Communication Review 25(4)*, 359-409.

McKnight, C. C. (In press). Critical evaluation of quantitative arguments. In G. Kulm (Ed.), *Assessing higher-order thinking in mathematics.* Washington, DC: American Association for the Advancement of Science.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: National Council of Teachers of Mathematics.

Peterson, L. V., & Schramm, W. (1954). How accurately are different kinds of graphs read?

Audio-Visual Communication Review 2(3), 178-189.

Siegler, R. S. (1978). The origins of scientific reasoning. In R. S. Siegler (Ed.), *Childrens' thinking: What develops?* Hillsdale, NJ: Lawrence Erlbaum Associates.

Tufte, E. R. (1983). *The visual display of quantitative information.* Cheshire, CN: Graphics Press.

Vernon, M. D. (1946). Learning from graphical material. *British Journal of Psychology 36(3)*, 145-157.

Vernon, M. D. (1950). The visual presentation of factual data. *British Journal of Educational Psychology 20(3)*, 174-185.

Washburne, J. N. (1927). An experimental study of various graphic, tabular and textual methods of presenting quantitative material. *Journal of Educational Psychology 18(6)*, 361-376.

Table 1. Three Questions for Graphical Data Processing

Circle answers to multiple choice questions. Otherwise, write where directed.

1. Approximately how many (age-adjusted) deaths from breast cancer per 100,000 population did Hong Kong have according to the data presented in this graph?
   a. About 10 deaths
   b. About 1,000,000 deaths
   c. About 70 deaths
   d. About 7,000,000 deaths
   e. Cannot tell from this graph.

2. Considering this graph only as dots on a piece of paper, what pattern do these dots (points) on the graph reveal?
   a. For the most part, as one considers larger and larger values of one variable, the other variable reveals increasingly larger values paired with the increasingly larger values of the second.
   b. For the most part, as one considers larger and larger values of one variable, the other variable reveals increasingly smaller values paired with the increasingly larger values of the second.
   c. For the most part, there is no pattern indicating a relationship between the two values that determine where each dot is located.
   d. There is a relationship other than that in "a" or "b" above. Please specify: [space for open-ended response]

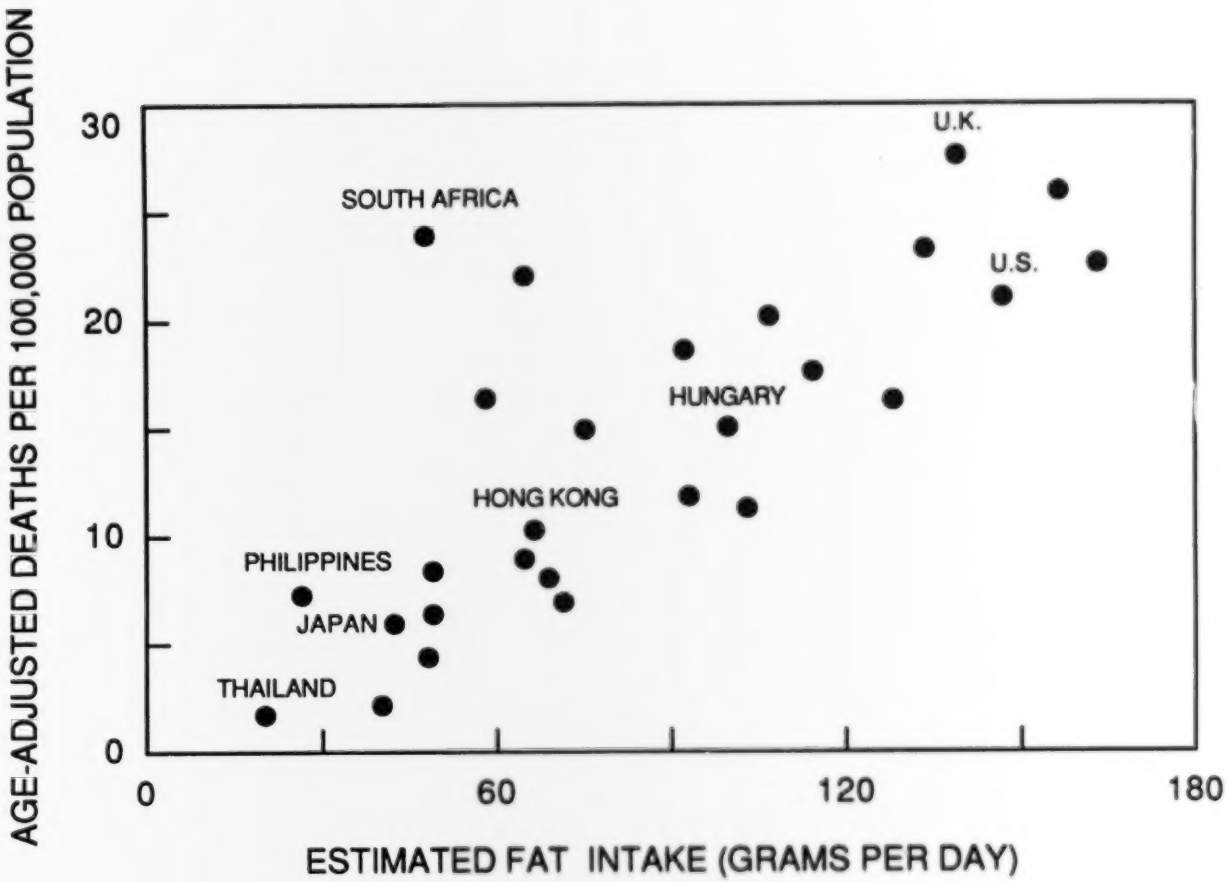
3. Given your understanding of the pattern among the dots in the graph and given your understanding of what each dot represents, state in your own words in *one or two sentences* what fact is represented by the relationship among the dots.

Note: In the instrument used to gather data, these three questions on this graph were followed by two others, one which involved assessing the evidential value of the graph for the proposition, "High intake of dietary fats help cause breast cancer," and the second that explored the bases on which individual decisions were made about evidential value.

Table 2. Responses to Question 3

1. As fat intake increases so do the number of deaths due to cancer.
2. The estimated fat intake produces more deaths.
3. The dots show how many deaths occur per 100,000 population, in relation to fat intake.
4. There really doesn't appear to be any relationship between the pattern of dots.
5. The more fat intake per day of consume *[sic]*, the higher your chances are of dying.
6. That the intake of fat per day increases the dot will be higher up on the age-adjusted deaths side.
7. The underprivileged nations of the world (listed at the bottom corner) have fewer fat people and age related deaths then *[sic]* the more urban and economical purposer *[sic; prosperous?]* nation *[sic]* of the world.
8. The dots on bottom of graph are third world countries, and the higher dots are mostly European countries. European countries take in more fats.
9. The graph indicates that as fat intake per capita increased the mortality rate in the individual populations from breast cancer increased as well.
10. In each country fat intake increases the cancer death rates.

FIGURE 1


AGE-ADJUSTED DEATHS PER 100,000 POPULATION
30
20
10
0
0
60
120
180
ESTIMATED FAT INTAKE (GRAMS PER DAY)
U.K.
SOUTH AFRICA
U.S.
HUNGARY
HONG KONG
PHILIPPINES
JAPAN
THAILAND